

University of Groningen

Reciprocity and Educational Evaluations by European Inspectorates

Leeuw, Frans L.

Published in:
Quality in Higher Education,

DOI:
[10.1080/1353832022000004331](https://doi.org/10.1080/1353832022000004331)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2002

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Leeuw, F. L. (2002). Reciprocity and Educational Evaluations by European Inspectorates: assumptions and reality checks. *Quality in Higher Education*, 8(2), 137. <https://doi.org/10.1080/1353832022000004331>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Reciprocity and Educational Evaluations by European Inspectorates: assumptions and reality checks¹

FRANS L. LEEUW

Evaluation Studies, Inspectie van het Onderwijs, Park Voorn 4, 3500 GS Utrecht, Holland

ABSTRACT *Many European countries have inspectorates of education and although they differ in some ways, all focus on the quality of education, all undertake evaluations and all strive for improvement in education. First, it will be argued that reciprocity between inspectors and inspectees (such as schools, colleges and institutes for vocational education) is important both for the evaluative work of inspectorates and for their work on behalf of quality improvement. Insights from the social, behavioural and economic sciences are used to underpin this point. Nevertheless, in practice it seems that only a minority of the 14 European inspectorates examined are involved in a reciprocal relationship with their evaluands/inspectees. Second, reciprocity and quality assurance organisations in higher education are discussed. It is argued that also in this field reciprocity between evaluator and evaluand is important. Third, several suggestions are made about how educational inspectorates can become more involved in reciprocal relationships without 'negotiating the truth'. Suggestions for further research are offered.*

Introduction

The assumptions underlying the activities of inspectorates of education usually include:

- whether or not quality control systems are effective and lead to a higher quality in education;
- the time frame of the activities;
- the way in which the evaluation and assessment activities are organised, including the level of independence of the position of the inspectorate from (central) government;
- the methodology of inspections;
- the contribution of the inspectorate's activities to educational innovations;
- the occurrence of unintended side-effects of quality evaluation activities, such as 'teaching to the test' and 'tunnel vision'.

Attention is often not paid to the issue of reciprocity between organisations evaluating the quality of education on the one side and schools and other education institutions on the other.

Reciprocity has two dimensions. The first dimension is the balance that is achieved between what information and data inspectorates and similar organisations want from the

educational field and what the field gets back in return from these activities. This is the *give-and-take-dimension* of reciprocity.

Reciprocity also concerns the balance between the evaluability (or transparency) that inspectorates request from institutions versus the evaluability/transparency of their own organisation and work. This is *you-too-me-too-dimension* of reciprocity.

The question to be answered in this paper is: to what extent are educational inspectorates in Europe engaged in operations that are focused on realising reciprocity between the evaluator (or: inspector) and evaluand (or: inspectee)?

Evaluators' or inspectors' organisations are the offices and persons active in the field of educational evaluations like Ofsted, the Inspection Générale de l'Éducation Nationale and the (Netherlands) Inspectie van het Onderwijs.

Evaluands are schools for primary and secondary education, colleges and institutions for vocational education and teacher training departments in universities. They are the recipients of the results from educational evaluations.

Structure of the Paper

First, the paper outlines why reciprocity is an important dimension when studying educational inspectorates. Next, several assumptions are put forward dealing with the relationship between reciprocity and evaluation. These assumptions can also be seen as 'issues' that should be taken into account if one studies inspectorates. Then some empirical evidence on the way in which reciprocity is dealt with by these offices in Europe is presented. Although the focus is on inspectorates that evaluate the quality of education in schools for primary, secondary and vocational/adult education, the issues raised are also relevant for the field of higher education. In particular, there is a reciprocity issue with regard to quality evaluation and assurance in higher education. Finally, some conclusions are drawn.

Why Is Reciprocity Important for Understanding the Relationship between Evaluators and Evaluands in the Field of Education?

Insights from Stakeholder Evaluations

In textbooks on evaluation, room is given to the role stakeholders play in carrying out evaluations. Rossi *et al.* (1999, p. 400) say that 'in undertaking their studies, evaluators usually find a diversity of individuals and groups with interest in their work and its outcomes'. Reineke (1991) has produced a number of suggestions about how to involve stakeholders in evaluations. The suggestions are based on the experiences working with school district staff in the US. Amongst others he suggests identifying stakeholders, to involve them early in evaluations and to do that continuously and actively. He also suggests creating a structure for involvement. 'Develop and use a conceptual framework based in content familiar to stakeholders, that can help keep dialogue focused'. Greene (1988) goes further in her participatory evaluation. She directly involves stakeholders in planning, conducting and analysing the data in collaboration with the evaluator, whose function might range from a team leader or consultant to that of a resource person to be called only as needed (Rossi *et al.*, 1999, pp. 57–58). Although this approach is not without critique (Pawson & Tilley, 1997), recent (comparative) studies of the role of inspectorates say little about these developments, nor about the concept of reciprocity. In the SICI-study by Standaert *et al.* (2000) the reciprocity concept is referred to only incidentally without any

analysis of its importance. However, in (quasi-popular) magazines like the *Times Higher Education Supplement* the lack of reciprocity that is to be attributed to, for example, Ofsted or the Quality Assurance Agency in the UK, is not only rather high, but also of concern to everybody interested in the impact and acceptability of the work of these organisations.

Insights from the Social and Economic Sciences

Social and behavioural scientists pay attention to reciprocity in the context of the social exchange theory (Blau, 1964; Homans, 1974; Klein Ikkink, 2000). Gouldner (1960) defines reciprocity as the ratio of giving and receiving support within a certain period of time. A relationship is reciprocal if the giving and receiving of support are balanced. For support, one may read many different things: financial, social, economic or 'information-oriented' support ('knowledge support').

Research focuses on the short versus long time perspective of reciprocity, indicating that the more reciprocal a relationship is the longer it will survive. 'The continuation of a relationship is (also) evaluated on the basis of previous investments' (Klein Ikkink, 2000, p. 3). Reciprocity helps to build social capital between people and organisations. Social capital concerns the importance of resources that, although possessed by other persons, are available to a given individual (or organisation) through his or her social relations to these other persons. The core of the theory and research of social capital is relatively simple. 'First, people better equipped with social resources—in the sense of their social network and the resources of others they can call upon—will succeed better in attaining their goals. Second, people will invest in relations with others in view of the perceived future value of the social resources made available by these relations' (Flap *et al.*, 1999). Third, organisations with more social capital than others will probably be better off. The central idea here is that more social capital is invested in reciprocal relationships.

As a review of more than 20 years of social science research in the field of social networks and social capital makes clear, person-to-person and organisational relationships with a high degree of social capital are often more effective and efficient than other types of relationships. This finding is important for anyone interested in the relationship between evaluator and evaluand (Bulder *et al.*, 1996; Flap *et al.* 1999; OECD, 2001). In line with this, Coleman (1990a) made the important statement that a cohesive relationship or network is a resource to its members because it promotes the willingness to co-operate with, and to provide help to, others. Mayne *et al.* (in press) suggest that the more social capital is available between organisations, the larger the likelihood is that partnership activities, for example, between public schools and private organisations, will be effective. Given the increased importance of collaboration between formal and informal organisations in the field of education (central government, local or regional government, quangos, grass roots organisations of parents and other stakeholders, 'students as consumers' and their organisations), this highlights the importance of reciprocity as a dimension relevant for the work of inspectors and evaluators.

Social science theory and research also helps to enlighten what happens to organisations that evaluate, assess, inspect and monitor when there is no, or only very limited, attention paid to reciprocity between evaluand and evaluator, or when the evaluation and monitoring activities are done in a rather formal, systematised way that only marginally involves relevant parties and stakeholders.

Transaction-costs economics, with its focus on the principal-agent relationship and the rational choice decision making by agent and principal, is, as Goshal and Moran (1996, p. 27) call it, 'bad for practice'. They have applied the sociological idea of a self-fulfilling

prophecy by predicting that systematised, formal controls, that influence trust within an organisation, may end up producing the opposite of what they want to accomplish. Goshal and Moran argue as follows:

- when organisations implement formal controls, they will become dependent on the information about compliance with standards and performance, produced by these systems;
- this shifts voluntary compliance and extra role-behaviour to compulsory compliance and work-to-rule activities of the evaluands;
- it encourages more difficult-to-detect opportunist behaviour by the evaluands.

The costs of removing these controls will grow until it is no longer an option for the organisation. Management's options for responding to opportunistic behaviour will narrow to one or more controls that would serve only to increase opportunistic behaviour. As this self-fulfilling prophecy plays itself out, management perceptions that employees are opportunistic would become increasingly valid.

One unintended consequence of this self-fulfilling prophecy is the increase in governance costs, making these organisations progressively uncompetitive. After all, the task of designing and implementing such controls is among the main causes for the build up of 'unneeded bureaucrats and wasteful bureaucratic practices' (Ghoshal & Moran, 1996). It can also enhance risk-averse behaviour, adversely affecting long-term performance.

In summary, if there is a focus on formalised monitoring and evaluation and a lack of reciprocity, this could turn monitoring and evaluation into a 'trust killer' (Leeuw, 2000).

Somewhat linked to this phenomenon is the idea that organisations that focus on performance monitoring may get involved in the performance paradox. This refers to a weak correlation between performance indicators and performance itself (Meyer & O'Shaugnessy, 1993; Meyer & Gupta, 1994). For 'performance' one can read 'quality'. The 'quality paradox' then concerns the weak correlation between quality evaluation and assessment activities on the one side and the quality itself on the other side.

This paradoxical phenomenon is caused by the tendency of performance indicators to run down over time. They lose their value as measurements of performance and can no longer discriminate between good and bad performers. As a result, the relationship between *actual* and *reported* performance or quality deteriorates. Deterioration of performance indicators is caused by four processes (Meyer & Gupta, 1994, pp. 330–342). The first process is called positive learning, that is, as performance or quality improves indicators lose their sensitivity in detecting bad performance or low quality. Second, perverse learning may occur, that is, when, although performance seems to improve, there is no actual improvement and perhaps even a deterioration of performance or quality. The third process, selection, refers to the replacement of poor performers with better performers, which reduces differences in performance. Fourth, suppression occurs when differences in performance are ignored.

Although the above focuses on unintended and even negative effects of auditing, evaluation and monitoring for trust, two things have to be stressed. One is that, given the amounts of (public) money spent on education, the need for transparency through monitoring and evaluation is evident and crucial in contemporary society. Second, probably not all 'types' of monitoring, auditing and evaluation are equally vulnerable to phenomena such as the 'performance paradox'. However, empirical research linking 'types' of monitoring and evaluation to this phenomenon under different institutional conditions does not yet appear to be available for the education field.

Assumptions on Reciprocity and Evaluating Educational Quality

Why is reciprocity important when doing educational evaluations by inspectorates? Realist evaluation methodology (Pawson & Tilley, 1997; Leeuw *et al.*, 1998) indicates that it is important to articulate the social and behavioural assumptions underlying reciprocity.

The first assumption is that the level of trust between evaluator and evaluand increases when inspectorates take reciprocity into account and act accordingly. Reciprocity implies dialogue, debate, openness and (intellectual) investor–investee relationships instead of primarily a top–down approach (Leeuw *et al.*, 1998; Power, 1999, p. 134ff). Trust is particularly important because, as Power (1999, p. 134) notes, ‘the audit society ... reflects a tendency not to trust’.

The more such a relationship is characterised by trust, the larger the probability that the evaluator takes into account the topics of educational quality, school effectiveness and school efficiency that really matter for schools, students, parents and teachers instead of taking only those into account that are political, administrative or procedural in nature. Power (1999, p. 131), citing Braithwaite, says:

Government inspectors ensure the quality of your records, not the quality of your deeds. If and when inspectors develop a reciprocal relationship with their inspectees, this role-behaviour probably will not occur any longer.

The more the relationship between inspector and inspectee is characterised by trust, the larger the probability that the inspectees will act upon the findings, evaluations and recommendations of the educational evaluation officers. Here the mechanism is that trust and mutual understanding act as an incentive for listening to the evaluator and not only listening because one is obliged to do so.

A fourth assumption is that the more reciprocity is a leading item on the agenda of inspectorates, the larger the probability is that they apply relevant and acceptable norms and standards. This is at least partly caused by the fact that the information exchange between evaluator and evaluand is more open and direct, when reciprocity is part of the culture of a review office.

The more reciprocity is a leading item on the agenda of educational evaluators, the longer the period will be that evaluands are positive about the contributions of the evaluators. Here the (behavioural) mechanism is that reciprocity creates a willingness to continue the relationship between both parties over time.

The more an inspectorate itself is open to scrutiny and willing to live up to the standards of transparency required from their evaluands, the more the evaluands will trust the office and subsequently act upon the recommendations, findings and evaluations of the evaluators.

The more an inspectorate puts reciprocity high on the agenda, the larger the probability is that the organisation will apply ‘proportional evaluations’, that is, will apply a practice in which the evaluator primarily uses the self-assessments or self-evaluations by the evaluand and only gets involved in on-site and primary data collection, when it is strictly needed. The mechanism here is that evaluators are concerned about the ‘behavioural costs’ (HEFCE, 2000) or ‘administrative burden’ their work can create for the evaluand and also strive to reduce these costs through proportionality in evaluation.

The more an inspectorate has reciprocity as a leading item on its agenda, the smaller the probability of a performance paradox occurring. Here the mechanism is first that due to the reciprocal relationship the evaluator has a better, ‘richer’ knowledge of what is going on inside the institution (and hence is better equipped to detect under- or over-representation

as a dimension of the performance paradox). Second, the trust mechanism works, because reciprocity will reduce the chances of cheating. What also might happen is that evaluands become more positive with regard to information-collecting approaches by inspectorates.

Reciprocity and the Evaluation of Educational Quality in Practice

Database and the Operationalisation of Concepts

Two sources have been used to find out to what extent reciprocity, in practice, is on the agenda of education inspectorates in Europe. The first approach uses a report edited by Maes *et al.* (1999); this report summarises characteristics of 14 inspectorates in Europe; it looks into most or all levels of education. As this report was published in 1999, the data collected are approximately 4 years old. As positions and activities of inspectorates, in recent years, have changed quite often, the problem with this database is that, according to some experts, several case-descriptions are no longer completely valid. Therefore, a second more informal resource has been used in which material obtained from an expert is central.²

The two dimensions of 'reciprocity' have been operationalised as follows. The nature of the database prevented the utilisation of more sophisticated operationalisations.

Operationalisation 1. This concerns the *you-too-me-too-dimension* of reciprocity. It focuses on the evaluability that characterises inspectorates or evaluation offices. This dimension is operationalised by information from the SICI-expert and by checking the content of Maes *et al.* (1999) for answers to the following question: *is the inspectorate/education review office evaluated externally and independently?*

The rationale behind this item is that the more inspectorates are evaluated externally and independently, the more they are open to critique and independent assessment and the more they show a willingness to be discussed 'in public'. The larger this willingness, the larger their contribution to (knowledge and experience) exchanges between inspectee and inspector.

Operationalisation 2. This focuses on the *give-and-take-dimension* of reciprocity. It concerns the extent to which the inspectorate is involving the public, schools, teachers, students and others in the process of finding and developing inspection norms and criteria. It also concerns the role the inspectorates play in informing the public and in participating in public debates on education and in sharing their knowledge with others. This dimension is operationalised for answers to the following two questions.

1. Are the evaluation norms and criteria of the inspectorate developed jointly with the educational field (that is inspectees) or are they developed by the quality evaluation organisation on its own?
2. To what extent is the inspectorate participating in public debates about education?

The rationale is that the more organisations participate in debates, and the more they involve inspectees in developing norms and criteria, the more reciprocal the relationship will be.

For six out of the 14 inspectorates, external evaluations by independent agencies are part of the game (Sweden, UK, Netherlands, Ireland, Northern Ireland and Scotland) (Table 1). Sometimes these external agencies are national audit offices, sometimes they are universities. Eight organisations are not evaluated by independent external actors.

With regard to the 'you-too-me-too-dimension' of reciprocity, this finding makes clear that less than half the inspectorates can be characterised as organisations that are themselves scrutinised in a public and external way.

Three out of the 14 organisations (Sweden, Netherlands, UK) develop norms and criteria for inspection and evaluation together with the educational field (that is the inspectees). Four inspectorates, to a lesser extent, involve schools, experts and other actors as partners in this process (Scotland, Portugal, Ireland, Flanders). The remaining seven inspectorates develop their norms, indicators and criteria without (bottom-up) participation from stakeholders.

The data show that almost all of the inspectorates participate in public debates about education.

The conclusion is that, in practice, reciprocity between evaluator and evaluand is not yet fully developed within the 14 national or regional education review offices. Only a minority of the European inspectorates of education is involved in a reciprocal relationship with their evaluands. If one compares this level of attention with the importance attached to the involvement of stakeholders, as is articulated in textbooks on evaluation, there is room for improvement for inspectorates. How to improve will be discussed in the final section but first attention will be paid to the issue of reciprocity when evaluating higher education programmes.

Reciprocity and Quality Evaluation in Higher Education

Harvey (1999) recently described the general way in which (institutional) evaluations within higher education take place in most of the (western) world:

Despite the very varied objects of evaluation and the array of different types of agencies [in many different countries], there is a surprising conformance in the methods that are adopted. Approaches to evaluation in higher education, as has frequently been pointed out, are heavily dependent on three basic elements:

- self-assessment (or submission);
- peer evaluation;
- statistical or performance indicators.

The results are prepared as a report that usually becomes a public document, albeit that a more detailed version may remain confidential. Typically, the procedure is for the institution or programme of study (or subject area) to produce a self-evaluation report or some other form of submission for assessment, such as a research profile. The qualitative self-evaluation is often complemented by statistical data.

The report, and the appropriate statistical data, are scrutinised by an external body. Sometimes more information is requested, either by the co-ordinating body or the team of 'respected' peers who will subsequently visit. This additional material may be received in advance or be available during the visit.

The peer-review panel visits the institution. Usually such a visit lasts between 1 and 4 days. They attempt to relate the self-assessment document to what they see or, in practice, hear. Often, they see relatively little as they spend most time closeted in a room having discussions with group after group of 'selected' discussants. In some cases, the peers may observe facilities or even the teaching and learning process itself, although the latter is rare (Harvey, 1999).

Quite often there are systems of follow-up implemented to check for the sustainability of the impact of the reports produced by quality assurance organisations.

TABLE 1. Content-analysis of the SICI-report on *Inspectorates of Education in Europe: A descriptive study* (Maes *et al.*, 1999) and information obtained from a SICI-expert

Country	Is the inspectorate/education review office evaluated (externally)?	Are the norms and criteria of the inspectorate/review office developed together with the educational field (i.e. inspectees/evaluands) or are they developed by the inspector's organisation on its own?	To what extent is the inspectorate/review office participating in public debates about education, sharing knowledge and evidence?
Austria	No	The process of finding norms and criteria that the inspectorate uses in its evaluation & oversight work are not developed together with inspectees/evaluands	Yes
The Czech Republic	No n n n n	The process of finding norms and criteria that the inspectorate uses in its evaluation & oversight work are not developed together with the inspectees/evaluands	Yes, but only when the organisation is asked to participate
England	Yes	Yes. Discussions with groups and organisations interested in education take place but Ofsted itself decides what is being evaluated (and how) and what not	Yes
Flanders	No	Somewhat	Yes, but only the minister who is responsible, acts
France	No	No, norms are not developed jointly with the inspectees/evaluands	Yes
The French-speaking community of Belgium	No	No, norms are not developed together with inspectees/evaluands	Yes, but very discretely and focusing on defending the point of view of the Minister
Hessen	No	No	Yes
Ireland	Somewhat	Somewhat	Yes, but not independently of the Dept of Education
The Netherlands	Yes	Yes	Yes
Northern Ireland	Yes	No, norms are not developed together with inspectees/evaluands	Yes, but in a low profile (p. 346)
North-Rhine-Westfalia	No	No, norms are not developed together with inspectees/evaluands	Occasionally
Portugal	No	Somewhat	No
Sweden	Yes	Yes	Yes
Scotland	Yes	Somewhat	Yes

For more detail, see Rohde and Leeuw (2001).

Usually the functions of quality assessment can be distinguished between accountability-oriented quality evaluation and improvement-oriented systems (Kalkwijk, 1998) or between control and support (Elmgren *et al.*, 1999). The organisations that carry out these tasks are largely funded by central governments but are at arms' length from the government. They are hardly ever part of inspectorates for education like the ones described in this paper. Usually they are quangos or equivalents. Their focus is on establishing the quality of higher education programmes, partly by reviewing the work of internal evaluators and peer-reviewers or 'visitation groups'. Sometimes quality assurance organisations carry out institutional evaluations. In a growing number of (European) countries, these organisations are also involved in accreditation.

Based on the analysis of the ways in which quality evaluations in higher education are organised in six countries by Scheele (1998) and the Center for Quality Assurance and Evaluations of Higher Education (CQAEHE) (1998) (Denmark, UK, Finland, France, Germany and Sweden), the following conclusions can be drawn with regard to the attention paid to reciprocity within these systems.

Three of these organisations (Centeret, QAA, Germany) are being evaluated externally and independently on the you-too-me-too-dimension. On the *give-and-take-dimension*, most of the organisations develop their norms and criteria together with the evaluands. However, the UK QAA is criticised for the 'heavy-handedness' of their approach (Turner, 2001). All six organisations are playing a role in the contribution to public debates on higher education.³

Reciprocity, therefore, is indeed on the agenda of these six organisations. However, Dill (1997, p. 38) has argued that nowadays the quality evaluation (systems) in higher education 'have been influenced by traditional theories about regulation, including the logics of hierarchical control through government inspection and performance auditing and reliance on enhancing market competition through deregulation and information provision'. Next to 'markets' and 'hierarchies' the role of 'social capital' has to be prominent. Social capital is 'the horizontal networks of social interaction that facilitate co-ordination, communication, the sharing of community norms and which promote individual engagement in a community'. Social capital, reciprocity and trust are closely related. Flap *et al.* (1999) and Putnam (2000) have shown how important social capital is for the productivity of organisations, including the quality of their services. However, Dill and Massy (1994) argue that current quality evaluation policies in higher education in fact can contribute to *reducing* social capital in academia, leading to what has been called 'hollowed collegiality'. Dill *et al.* (1996, p. 20) add to this the notion that 'in the current market condition of higher education, there are few incentives for administrators or institutions to renew or increase their collegial quality-assurance mechanisms'. This might be modified as follows: given the evidence about ways in which quality assurance models in education are being imitated and standardised,⁴ there is doubt that collegial-driven models will be much less developed and diffused than market/hierarchy-driven models.

Reciprocity, therefore, is also of relevance for higher education evaluators. Where trust between teacher, researcher, student and management is crucial, quality evaluators that do not invest enough in reciprocity run the risk of becoming 'trust killers'. This, in particular, can be true if they focus too much on their 'own' norms and criteria without discussing them in depth with their evaluands. As higher education has as one of its goals to help produce social capital, the consequences of being a 'trust killer' are serious. Some of the debates in the UK on the QAA can be interpreted in these terms.

Second, evaluations can produce unintended side-effects, including 'behavioural costs' like game playing, orchestration and staff stress (HEFCE, 2000). Other examples are

drop-out problems in academia due to a too heavy focus on the performance of schools and students, teaching to the test and 'templatism' (= focusing on 'templates and benchmarking [leading to] "dumbing down" by defining subjects in terms of minimal achievements with the associated danger of teaching to the syllabus'. The more reciprocal the relationship between evaluands and evaluators in higher education is, the more these side-effects can be addressed properly.

How Important Should Reciprocity Be for Inspectorates of Education?

The first reason is that direct and early feedback from students, schools, parents, pupils, teachers and boards is important because nobody has an ultimate claim to truth. This point has been made by Fitz Gibbon in her responses to, among others, Ofsted in the UK (Fitz-Gibbon, 2000). Knowledge, in the end, is always conjectural, needs to be tested and can never be taken for granted. That implies a plea for participation of inspectees in the process of inspection and for independent evaluations of the work of inspectorates.

Second, every research or measurement 'act' with people runs the risk of realising unintended side-effects or artefacts. This is also true for the work of inspectors and other educational evaluators (Leeuw, 1996, 2000). De Wolf (2001) recently found, for the Netherlands, that there is a strong correlation between the attitude of a school's management about the inspectors' work during their assessments and investigations on the one side and the final evaluation outcome by the inspectorate of the quality of schools on the other side. Only 23% of the primary schools that were rated as having a low quality are positive about the inspectors' on-site investigation, while 46% of the schools that are rated as good are positive about the 'visit' of the inspectors to the school. Apparently, on-site investigations and attitudes of school management are correlated. Therefore, it is the more necessary to realise a dialogue with evaluands like school boards, teachers and other inspectees. The rationale is that this facilitates the sharing of information, perspectives, experiences and 'theories' between evaluators and evaluands; that probably will increase the relevance of the work of inspectorates.

Linked to this is the third reason. If inspectorates are not deliberately involving students, teachers, boards, parents and others in their processes of developing norms and criteria, these stakeholders might bite back. Examples can already be found in the mass media and elsewhere.⁵ Biting back reduces the credibility of the inspector's organisation and work. As losing one's credibility is one of the most serious caveats for any evaluator (Leeuw *et al.*, 1994), this is a real danger.

It is somewhat of a paradox that where education is essentially a collaborative activity for children and young adults for a large number of years, the attention paid by inspectors to collaboration, partnering, and 'joined-up' activities with inspectees is still rather limited. Maybe the adage here should be to teach as one preaches. Patton (2000) goes even further: he is of the opinion that every time an evaluator is doing his or her job, he should ask himself what the evaluand gets out of this activity and what he or she learns from it.

Finally, when there is a reciprocal relationship, this reduces the likelihood of strategic behaviour by inspectees, because when an evaluand acts in this way, he or she will (immediately) lose credibility as a trustworthy partner to the inspectorate.

Three suggestions are offered as to how inspectorates of education could become more involved in establishing reciprocal relationships.

First, it is important to realise that inspectorates are becoming more involved in *knowledge transfer* from the offices to the schools, teachers, management and students. Evaluation offices in education know a lot about schools, their works, their organisation,

their quality and their pros and cons. The suggestion is to invest more in sharing that knowledge with inspectees. An example can be found in the work of the Georgia Council for School Performance (a kind of inspectorate) in the US. The knowledge transfer work of this organisation was recently awarded a prize by the American Evaluation Association. Fitz-Gibbon (2000) follows a somewhat similar route in the UK. By 'transferring knowledge' from the inspectorate to society, reciprocity will increase.

Second, realise that inspectorates understand that reciprocity fits nicely with the network concept of governance. More and more network management, steering through 'social capital', is on the agenda of governments (Mayne *et al.*, in press). Inspectorates should invest in understanding these processes for their own assessments.

Third, although inspectorates consider that no reciprocity is bad for practice they should also note that too much reciprocity will kill the independence of inspectorates and probably will realise the 'capture' of the offices by stakeholders and may even lead to 'negotiating the truth'. Therefore, there is a need for a balancing act between independence and reciprocity.

Some Suggestions for Further Research

Neo-institutionalism makes clear that although the (social and behavioural) mechanisms inside organisations are crucial for understanding the way organisations and institutions function, it is also true that organisational contingencies, and regional and national differences are important. To understand what the relationships are between reciprocity, trust, 'behavioural costs' of educational evaluations on the one side and the quality of education on the other side, it is needed to take these contingencies into consideration.

Theory-based evaluation that tries to go beyond good method and techniques as a basis for understanding what works has been on the evaluation agenda for many years. Chen (1990) published his advocacy of a theory-based approach 12 years ago and others started writing on this subject already in the early 1970s. This movement has been re-energised by the recent emergence of 'realist' and 'theory of change' approaches. The theory-based evaluation movement has been given added impetus by the challenging nature of contemporary (educational) policies. These are often expressed through complex programmes which when evaluated do not always offer clear-cut results. This is also true for education. If we are to generalise at all from evaluation findings, data need to be interpreted carefully—within a theoretical framework.

As was mentioned in the Delors report (Delors *et al.*, 1996), next to the well-known functions of education (education how to know, how to do and how to be), the function of education how to collaborate and how to create 'social capital' is important. Seen from this perspective, the possibility that educational evaluations 'help' to reduce instead of increase social capital within schools and institutes needs to be taken seriously. Since Coleman (1990b), it is known that social capital of schools and parents/students is an important co-determinant of the performance of schools and students (Morgan & Sorensen, 1999). Therefore, the relationships between reciprocity, social capital and educational evaluations needs to be further researched.

Notes

- [1] An earlier version of this paper was given as a keynote for the session 'Reciprocity and interaction in quality development', EU Congress 'The Meaning of Quality in Education', Karlstadt, Sweden, 2–4 April 2001. The author thanks the reviewers for their comments and suggestions.

- [2] To prevent misunderstanding: the 14 countries/regions are not to be considered representative for what is going on in Europe in general, although amongst them Nordic, southern, eastern and Anglo-Saxon countries/regions are involved. Thanks to Mr Johan van Bruggen, Netherlands Education Inspectorate and senior official of SICI, for the observation and for providing the additional material
- [3] I thank Mr Ko Scheele for this information.
- [4] See Hansen and Borum (1999) for a discussion of standardisation and imitation within evaluation in general and evaluation of higher education in particular. Organisational imitation (or mimicry) is linked to the theory of (organisational) isomorphism.
- [5] See the *Times Higher Education Supplement* during the first part of 2001. Almost every week there was an article, 'letter to the editor' or a column talking about the perception within academia of the role the UK QAA played.

References

- BLAU, P.M., 1964, *Exchange and Power in Social Life* (New York, Harcourt, Brace & Javanovitch).
- BULDER, B., LEEUW, F. & FLAP, H., 1996, 'Networks and evaluating public sector reforms', *Evaluation, the International Journal of Theory, Research and Practice*, 2(3), pp. 261-276.
- CENTER FOR QUALITY ASSURANCE AND EVALUATION OF HIGHER EDUCATION (CQAEHE), 1998, *A Status Report* (Copenhagen, Center Publishing).
- CHEN, H., 1990, *Theory-driven Evaluations* (London, Sage).
- COLEMAN, J., 1990a, *Foundations of Social Theory* (Boston, Harvard Bellknap Press).
- COLEMAN, J., 1990b, *Equality and Achievement in Education* (Boulder, Westview Press).
- DELORS, J. et al., 1996, *Learning, the Treasure Within* (Paris, Unesco).
- DILL, D.D., 1997, 'Accreditation, assessment, anarchy? The evolution of academic quality assurance policies in the US', in BRENNAN, J. et al. (Eds.) *Standards and Quality in Higher Education* (London, Jessica Kingsley).
- DILL, D.D. et al., 1996, 'Accreditation and academic quality assurance: can we get there from here?', *Change*, September/October, pp. 17-24.
- ELMGREN, C. et al., 1999, *Support or Control... or Both: The role of a national agency* (Stockholm, National Agency for Higher Education).
- FITZ-GIBBON, C.T., 2000, 'Keynote on educational evaluations', *Fourth Conference of the European Evaluation Society*, Lausanne, 12-14 October. Available online at: <www.europeanevaluation.org>.
- FLAP, H., BULDER, B. & VOLKER, B., 1998 'Intra-organisational networks and performance: a review', *Computational and Mathematical Organisation Theory*, 4(2), pp. 109-149.
- GOSHAL, M. & MORAN, A., 1996, 'Bad for practice: a critique of transaction cost theory', *Academy of Management Review*, 21(1), pp. 13-47.
- GOULDNER, A., 1960, 'The norm of reciprocity', *American Sociological Review*, 25(1), pp. 161-178.
- GREENE, J.C., 1988, 'Stakeholder participation and utilization in program evaluation', *Evaluation Review*, 15(4), pp. 91-116.
- HANSEN, H.F. & BORUM, F., 1999, 'The construction and standardization of evaluation: the case of the Danish university sector', *Evaluation, the International Journal of Theory, Research and Practice*, 6(3), pp. 303-329.
- HARVEY, L., 1999, 'Evaluating the evaluators', opening keynote at the *Fifth International Network of Quality Assurance Agencies in Higher Education (INQAAHE) Conference*, Santiago, Chile, 2 May.
- HEFCE, 2000, *Better Accountability for Higher Education* (London, HEFCE).
- HOMANS, G.C., 1974, *Social Behavior: Its elementary forms* (New York, Harcourt, Brace & Javanovitch).
- KALKWIJK, J.P.Th., 1998, 'Dancing on the slack rope', in SCHEELE, J.P. et al. (Eds.) *To Be Continued ...* (Den Haag, Elsevier/de Tijl).
- KLEIN IKKINK, C.E., 2000, *If I Scratch Your Back... Reciprocity and social support exchanges in personal relationships of older adults*, PhD thesis, Vrij Universiteit Amsterdam (Amsterdam, Thesis Press).
- LEEUEW, F.L., 1996, 'Performance auditing, new public management and performance improvement: questions and answers', *Accounting, Auditing and Accountability Journal*, 9(2), pp. 92-102.
- LEEUEW, F.L. 2000, 'Unintended side effects of auditing', in: RAUB, W. & WESSIE, J. (Eds.) *Management of Durable Relationships* (Amsterdam, Thesis Press).
- LEEUEW, F.L., RIST, F.C. & SONNICHSEN, R.C., Eds., 1994, *Can Governments Learn: Comparative perspectives on evaluation and organisational learning* (New Brunswick, Transaction Books).
- LEEUEW, F.L. et al., 1998, 'Evaluating anti-corruption initiatives: underlying logic and mid-term impact of a World Bank Program', *Evaluation, the International Journal of Theory, Research and Practice*, 5, pp. 194-219.
- MAES, B. et al. 1999, *Inspectorates of Education in Europe: A descriptive study* (Brussels, SICI/DVO).

- MAYNE, J. *et al.*, in press, 'Networks and partnering arrangements in the collaborative state: new topics for evaluation and auditing', in GRAY, A. *et al.* (Eds.) *Sharing Power: Evaluation and management in the collaborative state* (New Brunswick, Transaction Books).
- MEYER, M. & GUPTA, V., 1994, 'The performance paradox', *Research in Organisational Behavior*, 16(1), pp. 309–369.
- MEYER, K. & O'SHAUGNESSY, K., 1993, 'Organisational design and the performance paradox', in SWEDBERG, R. (Ed.) *Explorations in Economic Sociology* (New York, Russell Sage Foundation).
- MORGAN, S. & SORENSSEN, A., 1999, 'Parental networks, social closure and mathematics learning: a test of Coleman's social capital explanation of school effects', *American Sociological Review*, 64(3), pp. 661–681.
- OECD, 2001, *The Well-being of Nations: The role of human and social capital* (Paris, OECD).
- PATTON, M., 2000, 'Utilization of evaluations', keynote address to the *Fourth Conference of the European Evaluation Society*, Lausanne, 12–14 October. Available online at: <www.europeanevaluation.org>.
- PAWSON, R. & TILLEY, N., 1997, *Realist Evaluations* (London, Sage).
- POWER, M., 1999, *The Audit Society: Rituals of verification* (Oxford, Oxford University Press).
- PUTNAM, R.D., 2000, *Bowling Alone: The collapse and revival of American community* (New York, Basic Books).
- REINEKE, R.A., 1991, 'Stakeholder involvement in evaluation: suggestions for practice', *Evaluation Practice*, 12(1), pp. 39–44.
- ROHDE, E. & LEEUW, F.L., 2001, *Reciprociteit en onderwijsinspecties: Appendix*, Internal Working Paper (Utrecht, Netherlands Education Review Office).
- ROSSI, P.H. *et al.*, 1999, *Evaluation: A systematic approach* (London, Sage).
- STANDAERT, R., 2000, *Inspectorates of Education in Europe: A critical analysis* (Utrecht/Brussels, SICI & Ministerie van de Vlaamse Gemeenschap).
- TURNER, D., 2001, 'The gentle touch', *Guardian*, 30 August.
- DE WOLF, I., 2001, *Resultatenvergelijking oordelen inspectie en attitudes schoolleiding*, Internal Working Paper (Utrecht, Netherlands Education Review Office).